DOCUMENT RESUME

ED 193 544 CG 014 695

AUTHOR Sauser, William I., Jr.
TITLE A Comparison of the Effects of Rater Training and Participation on Sources of Variance in a Set of BARS Ratings.
PUB DATE Mar 80
NOTE 17p.; Paper presented at the Annual Meeting of the Southeastern Psychological Association (26th, Washington, DC, March 26-29, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Behavior Rating Scales; Comparative Analysis; Participation; Psychometrics; *Student Evaluation of Teacher Performance; Student Teacher Relationship; Teacher Effectiveness; Teaching Styles; *Test Construction; *Test Validity; *Training
IDENTIFIERS *Variance (Statistical)

ABSTRACT

The effects of training and participation on sources of variance in a set of ratings of college classroom teaching effectiveness were compared. College students (N=96) were randomly assigned to four cells of the experimental design. Subjects in cells (A) and (B) participated in the construction of a set of behaviorally-anchored rating scales (BARS) of five aspects of college classroom teaching performance, while subjects in cells (C) and (D) performed a control task. Later, subjects in cells (A) and (C) were exposed to a rater training program, while subjects in cells (B) and (D) performed a control task. All subjects then evaluated five standardized simulated professors using the BARS. Training significantly reduced the overall elevation of the ratings; participation did not. Neither participation nor training significantly reduced the variance attributable to the category of behavior being evaluated. Both participation and training significantly reduced variance attributable to the professor being rated. Participation significantly increased the Category x Professor effect while training did not. There were no significant interactions among the treatments with regard to effects on any of the above characteristics of ratings. Findings suggest that, for these four characteristics of ratings, participation and training operate independently of each other. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193544

A Comparison of the Effects of Rater Training and Participation on Sources of Variance in a Set of BARS Ratings

William I. Sauser, Jr.

Auburn University

Paper presented at the meeting of the Southeastern Psychological Association

Washington, D.C.

March 1980

CG 014695

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William I. Sauser

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Purpose

The most popular criterion measure employed in personnel research and practice today is the rating scale (Blum & Naylor, 1968, pp. 197-198). Despite its popularity, the rating method has been severely criticized due to questionable levels of reliability and validity (Ronan & Schwartz, 1974) and susceptibility to "rating errors" such as leniency, central tendency, and halo (Smith, 1976). While several techniques have been used in attempts to improve the quality of ratings as criteria, the two approaches found generally most successful are rater training (Guilford, 1954, p. 280; Latham, Wexley, & Pursell, 1975) and rater participation in scale construction (Campbell, Dunnette, Arvey, & Hellervik, 1973; Smith & Kendall, 1963). The industrial/organizational psychology literature contains numerous studies of the effectiveness of these two approaches, yet direct comparisons of their effects on psychometric characteristics of ratings are scarce. The purpose of this study was to directly compare the effects of training and participation on sources of variance in a set of ratings of college classroom teaching effectiveness.

Method

Ninety-six undergraduate students taking courses in psychology at a large southeastern university were randomly assigned to four cells of the experimental design: (a) Both Participation and Training, (b) Participation Only, (c) Training Only, and (d) Neither Participation nor Training. Subjects in cells (a) and (b) participated in the construction of a set of behaviorally anchored rating scales (BARS) for measuring five aspects of college classroom teaching performance, while subjects in cells (c) and (d) performed a control task. Later, subjects in cells (a) and (c) were exposed to a rater training program, while subjects in cells (b) and (d) performed a control task. All subjects then evaluated five standardized simulated professors using the BARS. These "simulated professors" consisted of short biographical descriptions followed by behavioral diaries containing scaled incidents obtained during the BARS construction process.

Table 6. Definitions of the Five Categories of College Classroom Teaching Behavior

A. Relationships with Students. This category refers to the way the professor treats his/her students both in and out of class. It includes such things as talking with students before, during, and after class, interacting with and counseling students in the office and elsewhere regarding course-related and personal problems, knowing students' names, and treating students with respect in class.

B. Ability to Present the Material. This category refers to the way the professor organizes the material and presents it to the class. It includes such things as coming to class well-prepared and on time, organizing the material in a logical manner, speaking and writing clearly, and using examples, audio-visual aids, and other devices to get the material across to the students.

C. Interest in Course and Material. This category refers to the professor's knowledge of and interest in the material he/she is trying to teach. It includes such things as being able to answer questions and elaborate on the material, showing enthusiasm for the course, and reading and researching to keep current and learn more about the subject matter.

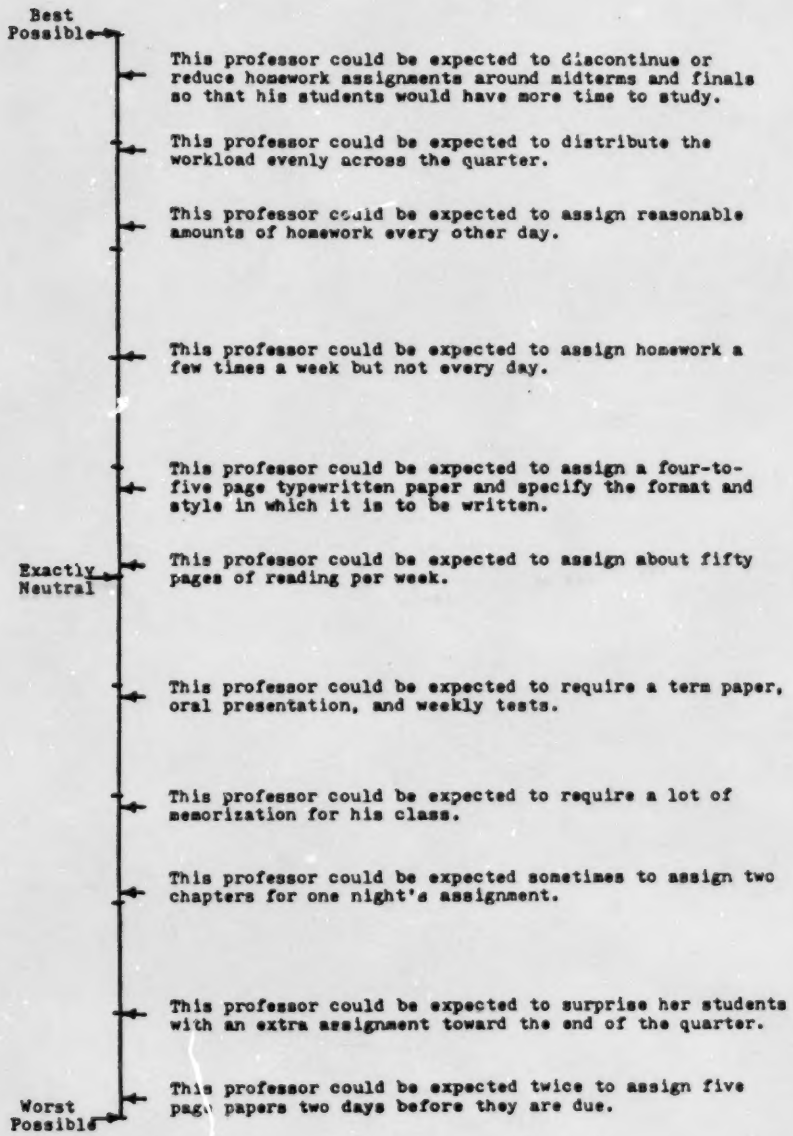
D. Reasonableness of the Workload. This category refers to the amount of work (reading, homework problems, class and lab work, papers, tests, etc.) assigned by the professor. It includes such things as clearly specifying assignments and due dates, scheduling the work evenly throughout the quarter, and keeping the workload appropriate to the credit-hour value of the course.

E. Fairness of Testing and Grading. This category refers to the fairness of the professor's testing and grading policies. It includes such things as stating how grades are to be determined, testing over appropriate material, and grading without bias.

APPENDIX F-1 (Cont'd)

D. Reasonableness of the Workload

This dimension refers to the amount of work (reading, homework problems, class and lab work, papers, tests, etc.) assigned by the professor. It includes such things as clearly specifying assignments and due dates, scheduling the work evenly throughout the quarter, and keeping the workload appropriate to the credit-hour value of the course.


Best Possible
This professor could be expected to discontinue or reduce homework assignments around midterms and finals so that his students would have more time to study.
This professor could be expected to distribute the workload evenly across the quarter.
This professor could be expected to assign reasonable amounts of homework every other day.
This professor could be expected to assign homework a few times a week but not every day.
This professor could be expected to assign a four-to-five page typewritten paper and specify the format and style in which it is to be written.
Exactly Neutral
This professor could be expected to assign about fifty pages of reading per week.
This professor could be expected to require a term paper, oral presentation, and weekly tests.
This professor could be expected to require a lot of memorization for his class.
This professor could be expected sometimes to assign two chapters for one night's assignment.
This professor could be expected to surprise her students with an extra assignment toward the end of the quarter.
This professor could be expected twice to assign five page papers two days before they are due.
Worst Possible

APPENDIX G-1

PROFESSOR L

Professor L is a 29-year-old male Assistant Professor who is new at Auburn. He has long red hair, a full beard and moustache, and is a heavy smoker. He usually wears jeans and flannel shirts, boots, and a black leather jacket to class. He is not very well known in his field but has initiated a number of research projects since arriving at Auburn. He teaches a 5-hour, 300-level science course with a laboratory.

You observed the following things about Professor L while taking his course:

He used a variety of methods to present the material, including films, tapes, and experiments.

He told the class he would grade on a 10-point scale, then actually used a 7-point scale to assign final grades.

He often described his own fascination with the material he was covering.

He gave a mid-term and final only.

He assigned only as much homework as was necessary to learn the material thoroughly.

He was attentive and helpful in class, but was generally unavailable for outside help.

He gave plenty of time to read the material and discussed it thoroughly in class.

Once when asked a question in class he lost patience with himself because he could not answer it.

He always left promptly after giving his lectures.

When asked by his students what to study for a test, he said, "I don't know, I haven't made it out yet."

He did not curve grades even if the average score was in the 50s or 60s.

He gave a student unclear and evasive answers to her questions when she visited his office.

His lectures were boring and unorganized.

He assigned about two hours worth of work to be done during his three-hour laboratory so that no one would have to rush.

APPENDIX G-1 (Cont'd)

He took his lectures straight from the book and never gave examples.

He often told the class about interesting articles he had read or experiments he had heard about.

Although he gave his office number and hours on the first day of class, he did not encourage the students to come see him.

Once when confounded by a student's question in class he spent several hours of his own time that afternoon researching material for an answer.

He reduced the workload at the end of the quarter when he realized that his students did not have enough time to complete all of the assignments.

He sought student input to support his conclusions in class.

Table 7. Scale Values in the Simulated Professor x Category Matrix

Category	Simulated Professor L	M	N	O	P	Row Sum	Row Mean	Row Variance
A	4.0	10.0	8.0	2.0	6.0	30.0	6.0	8.0
B	6.0	2.0	10.0	4.0	8.0	30.0	6.0	8.0
C	8.0	4.0	6.0	10.0	2.0	30.0	6.0	8.0
D	9.6	6.0	2.0	8.0	4.0	29.6	5.9	7.4
E	2.0	8.0	4.0	6.1	10.0	30.1	6.0	8.0
Column Sum	29.6	30.0	30.0	30.1	30.0			
Column Mean	5.9	6.0	6.0	6.0	6.0			
Column Variance	7.4	8.0	8.0	8.0	8.0			

Analysis

The data were analyzed in a split-plot factorial ANOVA with Participation and Training (two levels each) serving as between-subjects factors and Categories (of performance) and Professors (five levels each) as within-subjects factors. Additional analyses were performed to interpret various significant interactions among the factors. The omega-square statistic was employed to determine the practical significance of statistically significant effects.

Table 8. Study One ANOVA Table--All Subjects

Source	df	SS	F^a	ω^2
Participation	1	1.6964	0.77	-
Training	1	31.4683	14.35*	.0013
Part x Train	1	0.4991	0.23	-
Subjects w. groups	4	8.7720	0.63	-
Categories	4	314.3566	22.57***	.0145
Part x Cat	4	27.6427	1.99	-
Train x Cat	4	7.5794	0.54	-
Part x Train x Cat	4	17.2175	1.24	-
Cat x Subj w. grp	16	82.7864	1.49	-
Professors	4	67.1426	4.82***	.0026
Part x Prof	4	47.9253	3.44**	.0016
Train x Prof	4	34.0416	2.44*	.0010
Part x Train x Prof	4	20.5515	1.48	-
Prof x Subj w. grp	16	54.7207	0.98	-
Cat x Prof	16	12071.2852	216.72***	.5786
Part x Cat x Prof	16	102.4909	1.84*	.0023
Train x Cat x Prof	16	79.9003	1.43	-
Part x Train x Cat x Prof	16	26.9005	0.48	-
Cat x Prof x Subj w. grp	64	167.8245	0.75	-
Residual	2183	7599.6374	-	-
Total	2382	20764.4391	-	-

[a]All effects were tested against Residual except for Participation, Training, and Part x Train, which were tested against Subjects w. groups.

*$p < .05$

**$p < .01$

***$p < .001$

Table 9. Study One ANOVA Table--Participant Subjects Only

Source	df	SS	F^a	ω^2
Training	1	11.6662	3.21	-
Subjects w. groups	2	7.2767	1.18	-
Categories	4	156.9696	5.37*	.0121
Train x Cat	4	9.8173	0.34	-
Cat x Subj w. grp	8	58.4766	2.36*	.0032
Professors	4	25.5242	2.06	-
Train x Prof	4	47.1598	3.81**	.0033
Prof x Subj w. grp	8	14.4242	0.58	-
Cat x Prof	16	6661.7041	134.54***	.6279
Train x Cat x Prof	16	57.5041	1.16	-
Cat x Prof x Subj w. grp	32	122.3188	1.24	-
Residual	1084	3354.7129	-	-
Total	1183	10527.5546	-	-

[a]All effects were tested against Residual except for Categories and Train x Cat, which were tested against Cat x Subj w. grp; and Training, which was tested against Subjects w. groups.

*$p < .05$

**$p < .01$

***$p < .001$

Table 10. Study One ANOVA Table--Non-participant Subjects Only

Source	df	SS	F[a]	ω^2
Training	1	20.4132	27.30*	.0016
Subjects w. groups	2	1.4953	0.19	-
Categories	4	184.8966	11.97***	.0165
Train x Cat	4	15.4290	1.00	-
Cat x Subj w. grp	8	24.3098	0.79	-
Professors	4	89.2225	5.77***	.0072
Train x Prof	4	7.4333	0.48	-
Prof x Subj w. grp	8	40.2965	1.30	-
Cat x Prof	16	5511.2407	89.18***	.5322
Train x Cat x Prof	16	50.0515	0.81	-
Cat x Prof x Subj w. grp	32	45.5057	0.37	-
Residual	1099	4244.9245	-	-
Total	1198	10235.2187	-	-

[a]All effects were tested against Residual except for Training, which was tested against Subjects w. groups.

*$p < .05$

***$p \leq .001$

Table 11. Study One ANOVA Table--Trained Subjects Only

Source	df	SS	F^a	ω^2
Participation	1	2.1316	3.69	-
Subjects w. groups	2	1.1542	0.17	-
Categories	4	179.2890	5.74*	.0148
Part x Cat	4	34.3782	1.10	-
Cat x Subj w. grp	8	62.4300	2.27*	.0035
Professors	4	33.1986	2.42*	.0019
Part x Prof	4	50.7707	3.70**	.0037
Prof x Subj w. grp	8	35.1496	1.28	-
Cat x Prof	16	5706.6370	103.97***	.5643
Part x Cat x Prof	16	56.1690	1.02	-
Cat x Prof x Subj w. grp	32	100.8442	0.92	-
Residual	1093	3749.4051	-	-
Total	1192	10011.5572	-	-

[a]All effects were tested against Residual except for Categories and Part x Cat, which were tested against Cat x Subj w. grp.; and Participation, which was tested against Subjects w. groups.

*$p < .05$

**$p < .01$

***$p < .001$

Table 12. Study One ANOVA Table--Untrained Subjects Only

Source	df	SS	F^a	ω^2
Participation	1	0.1454	0.04	-
Subjects w. groups	2	7.6178	1.08	-
Categories	4	142.5991	10.09***	.0120
Part x Cat	4	10.9008	0.77	-
Cat x Subj w. grp	8	20.3564	0.72	-
Professors	4	68.0058	4.81***	.0050
Part x Prof	4	17.3102	1.23	-
Prof x Subj w. grp	8	19.5712	0.69	-
Cat x Prof	16	6444.4720	114.03***	.5956
Part x Cat x Prof	16	73.7224	1.30	-
Cat x Prof x Subj w. grp	32	66.9803	0.59	-
Residual	1090	3850.2323	-	-
Total	1189	10721.4136	-	-

[a]All effects were tested against Residual except for Participation, which was tested against Subjects w. groups.

***$p < .001$

Conclusions

The major findings were: (1) Training significantly reduced the overall elevation of the ratings, whereas participation did not. (2) Neither participation nor training significantly reduced the variance attributable to the category of behavior being evaluated. (3) Both participation and training significantly reduced variance attributable to the professor being rated. (4) Participation significantly increased the Category x Professor effect (discriminant validity) while training did not. (5) There were no significant interactions among the treatments with regard to effects on any of the above characteristics of ratings. Thus, it appears that participation and training operate independently of each other, at least as far as these four characteristics of ratings are concerned.

References

Blum, M.L., & Naylor, J.C. Industrial psychology: Its theoretical and social foundations. New York: Harper & Row, 1968.

Campbell, J.P., Dunnette, M.D., Arvey, R.D., & Hellervik, L.V. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.

Guilford, J.P. Psychometric methods (2nd ed.). New York: McGraw-Hill, 1954.

Latham, G.P., Wexley, K.N., & Pursell, E.D. Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 60, 550-555.

Ronan, W.W., & Schwartz, A.P. Ratings as performance criteria. International Review of Applied Psychology, 1974, 23, 71-82.

Smith, P.C. Behaviors, results, and organizational effectiveness: The problem of criteria. In M.D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Smith, P.C., & Kendall, L.M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.